Measures of Center and Spread

Mean

symbols: $\bar{x}$ or μ

Formula:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

n ← total # of obs.

x_i ← observations (data pts)

$\frac{1}{n}$ ← $\div n$

$x_1 + x_2 + x_3 + \ldots + x_n$

Median

Process:

* sort data
* middle #
  - odd # of data pts, M = center #
  - even # of data pts, M = avg of 2 center #'s

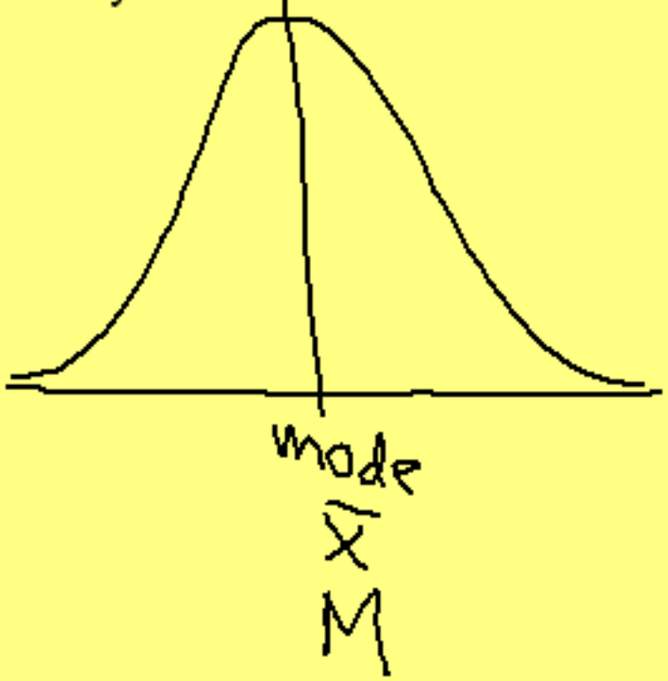
Mean	Median
• avg • non-resistant	• middle # • resistant ↳ not affected by outliers

$\bar{x} = 140$

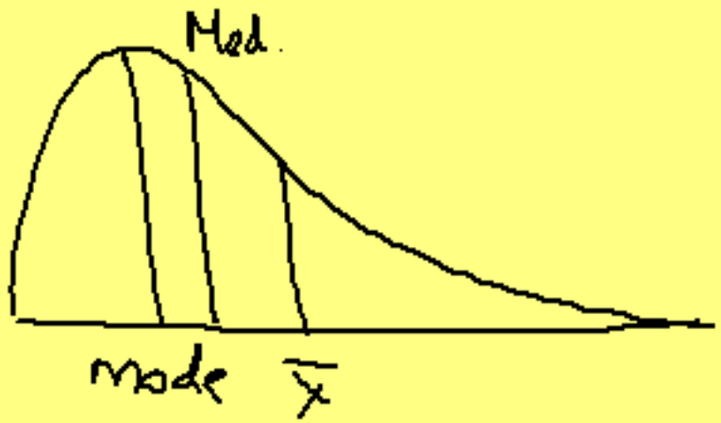
$M = 100$

Where are the measures of center for different shaped distributions?

Symmetric (roughly)

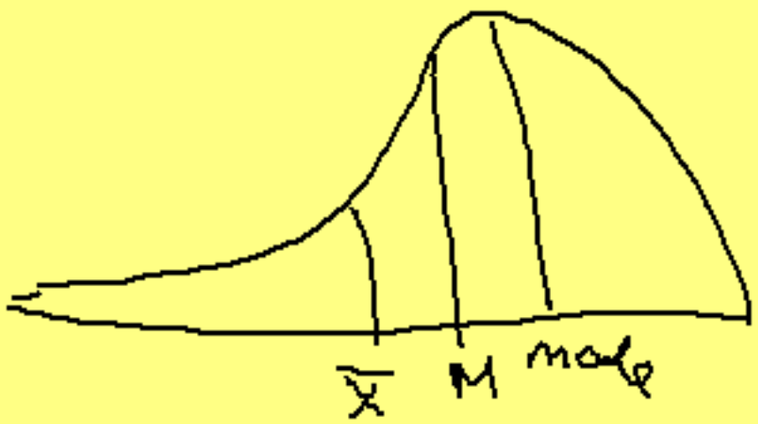


Right Skewed



$\bar{x} > M$

Left Skewed



$\bar{x} < M$

MEDIAN

Measure of Center: Median

Measure of Spread: 5 # summary

Min Q1 Med Q3 Max

Used for: skewed or symmetric

resistant to outliers

Range: full spread of data (a, b)

Quartiles: splits data into 4 equal sections
Q_3 = split the upper ½ of data
Q_1 = split the lower ½ of data

IQR:
Inter Quartile range $= Q_3 - Q_1 = \#$

- What is the IQR used for?
  - Shows... middle 50% of data
  - Helps... identify outliers

1.5x

IQR Test for possible outliers:

- Find IQR (#)
- $1.5 \times IQR$ (#)
- $Q_3 + (1.5 \times IQR) = c$

$Q_1 - (1.5 \times IQR) = d$

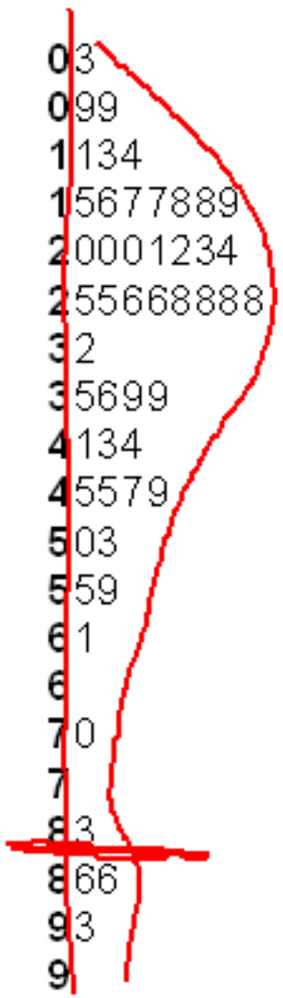
(d, c)

- Anything outside this range is considered an outlier

Why is it just for POSSIBLE outliers?

Example: Supermarket Spending- data below is the amount (rounded to the nearest $) spent in the supermarket. **Do a test for possible outliers**

Supermarket Spending



Med: 28

n=50

Q1 = 19

Q3 = 45

IQR = Q3 − Q1 = 26

1.5(26) = 39

Q1 − 39 = −20

Q3 + 39 = 84

(~~−20~~, 84)

* use picture & IQR test to determine outliers

Percentiles:

an obs that has a certain % of the data below it (or equal to)

Example: Use the data above to find the percentile for the observation $24.

$\frac{20}{50} = 40\%$

Special Percentiles:

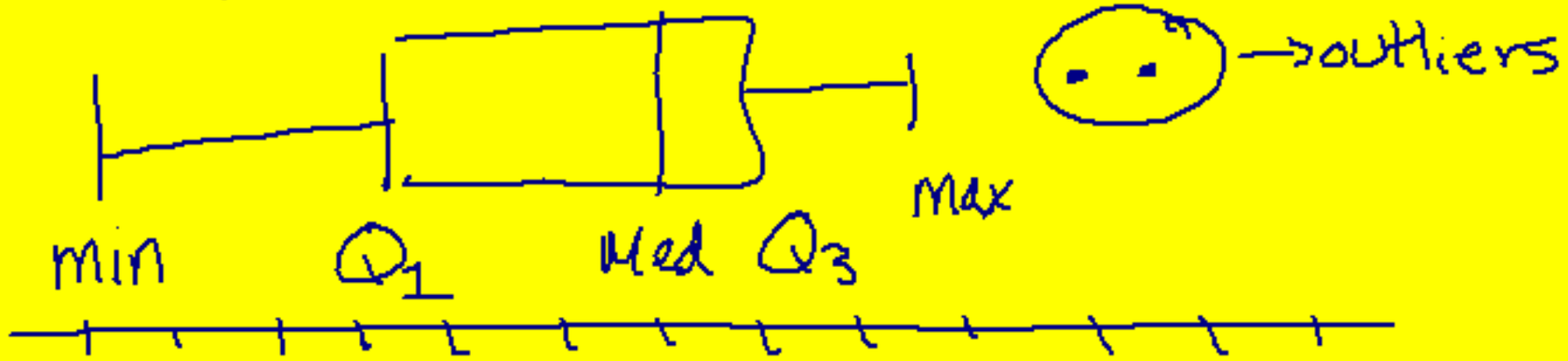
Q1 = 25th

Median = 50th

Q3 = 75th

Distribution (using Median):

Modified Boxplot



5# summary

Side-by-side

MEAN

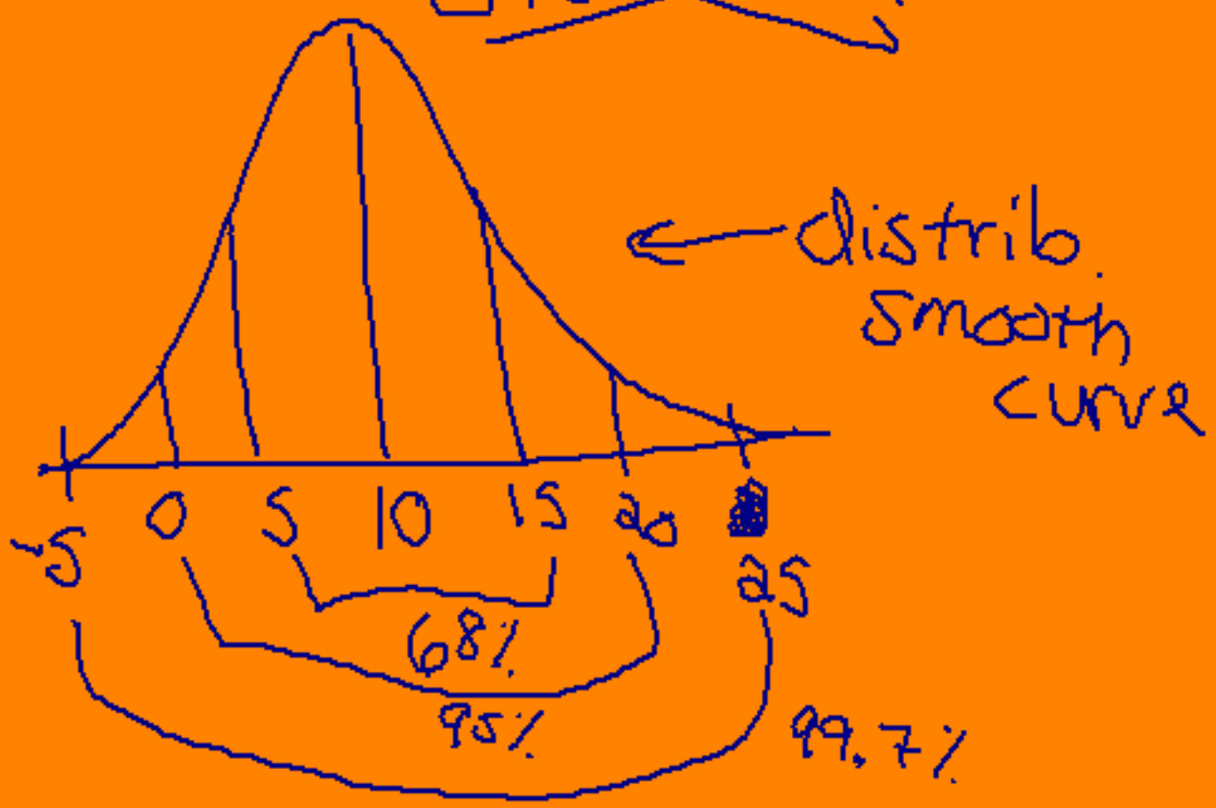
Measure of Center: Mean

Measure of Spread: standard deviation

Used for: symmetric

~~Skewed~~

Std. dev = 5



What IS Standard Deviation?

average distance that pts. are from their mean.

How do we calculate it?

Variance

Symbol: σ^2 or s^2

Formula: (Std. dev)^2

Std. Deviation

Symbol: σ (sigma), s

σ ← pop., s ← sample

Formula:

$$\sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$\bar{x}$ = sample

μ = pop.

- Note: n - 1 = degrees of freedom
- $x_i - \bar{x}$ = deviation

12

Questions: (p. 52-53 in book)

1. **Why do we square the deviations?**

– so all deviations are +

mean = center

$\sum$ + deviations =

$\sum$ – deviations

– 4

+ 1

+ 3

2. **Why do we divide by (n - 1?)**

n = 10

$\bar{x}$ = 20

– free

3. **Why use the std. deviation and not the variance?**

regular units

σ

squared units

σ^2

Notes on Std. Dev:

- Only use when... mean is measure of center
- S = 0 means... data is all same
- Unless there is no spread,... (s=0) S is a + #
- As observations become more spread out... S ↑
- NON-RESISTANT

Easy way to calculate Std. Deviation:

1-var stats

Please do worksheets:

1.2- Measures of Central Tendency

1.2- Measures of Spread

Answer keys will be left on the front table so you can check your work.